



**QUEEN'S
UNIVERSITY
BELFAST**

Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy

Jiménez-Carvelo, A. M., Osorio, M. T., Koidis, A., González-Casado, A., & Cuadros-Rodríguez, L. (2017). Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy. *LEBENSMITTEL-WISSENSCHAFT UND-TECHNOLOGIE-FOOD SCIENCE AND TECHNOLOGY*, 86, 174-184. <https://doi.org/10.1016/j.lwt.2017.07.050>

Published in:

LEBENSMITTEL-WISSENSCHAFT UND-TECHNOLOGIE-FOOD SCIENCE AND TECHNOLOGY

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2017 Elsevier.

This manuscript is distributed under a Creative Commons Attribution-NonCommercial-NoDerivs License

(<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

**CHEMOMETRIC CLASSIFICATION AND QUANTIFICATION OF OLIVE OIL
IN BLENDS WITH ANY EDIBLE VEGETABLE OILS USING FTIR-ATR AND
RAMAN SPECTROSCOPY**

Ana M. JIMÉNEZ-CARVELO¹ ✉, María Teresa OSORIO², Anastasios KOIDIS²,
Antonio¹ GONZÁLEZ-CASADO, Luis CUADROS-RODRÍGUEZ¹

¹ Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n. E-
18071 Granada, Spain.

² Institute for Global Food Security, Queen's University, 18-30 Malone Road, Belfast
BT9 5BN, Northern Ireland, UK

Keywords : vegetable oils, discrimination, fingerprinting, pattern recognition,
spectroscopic techniques

✉ Corresponding author: phone: +34 958240797; fax: +34 958243328; email:
amariajc@ugr.es

Abstract

Samples of olive oils (n=67) from different qualities and samples of other vegetable edible oils (including soybean, sunflower, rapeseed, corn oil etc; n=79) were used in this study as pure oils. Previous to spectroscopy analysis, a transesterification step was applied to the pure vegetable oil samples and all the different oil blends were then prepared to create in-house blended samples. Spectral acquisition was performed with typical parameters to collect the FTIR and Raman fingerprints. For the olive/non-olive classification model, three classification strategies have been applied: (i) one input-class (1iC) classification; (ii) two input-class (2iC) classification; and (iii) one input-class plus one 'dummy' class classification (or *pseudo* two input-class (*p2iC*) classification). The multivariate classification methods used were k-nearest neighbours (kNN), partial least squared-discriminant analysis (PLS-DA), one-class partial least squares (OCPLS), support vector machine classification (SVM-C), and soft independent modelling of class analogies (SIMCA). The multivariate quantification method used was partial least square-regression (PLS-R). FTIR fingerprints showed excellent classification ability to distinguish pure olive from non-olive oil. When PLS-DA or SVM-C techniques are applied, 100% of olive oil samples and 92% of other vegetable edible oils are correctly classified. In general FTIR fingerprints were more discriminative than Raman's in both classification and regression scenarios.

1. INTRODUCTION

As a natural product that is produced using 'only mechanical means' from olive drupes, olive oil is protected by various regulations and institutions such as the EU Regulations (Regulation UE, 2016; Regulation UE, 2011; Commission Regulation EEC, 2016) and Codex Alimentarius (Codex Stan, 2015). Due to its increasing popularity, it has always been the target for fraudulent practises such as substitution fraud with cheaper oils (blends). To prevent that, authenticity of olive oil is described adequately in the legislation. The top two qualities of olive oil that exist are the extra-virgin and the virgin olive oil and both of them must comply to certain well defined physical, chemical and sensorial parameters. There are several standard methods that are used to determine these parameters. For example, with the use of chromatographic techniques detection of several major and minor constituents of olive oil (fatty acids, tocopherols, carotenoids etc.) is achieved. Nowadays rapid and novel methods are continuously developed (such as those based on spectroscopy), as alternatives to the standard methods offering speed, efficiency (less resources required) and accuracy in authenticity testing.

Actually, studies about authentication of olive oil using spectroscopic techniques are based on the application of chemometric tools to develop multivariate models that are able to differentiate pure olive oils from adulterated olive oil with other vegetable edible oil. Then, the proportion of olive oil in these blends is quantified; therefore, although blends of olive oil with other vegetable oils are allowed by the legislation, there is a restriction of labelling them as "olive oils" if the olive oil in the blend does not exceed 50% (Regulation UE, 2016). Consequently, a proper method of control must be established. Sun, Lin, Li, Shen and Luo (2015) reported: (i) a principal component analysis (PCA) model to discriminate extra virgin olive oil from binary blends of olive oil with camellia oil, soybean oil, sunflower oil and corn oil; and (ii) a quantification model using partial least squares (PLS) to quantify the olive oil in binary blends. López-Díez

and Goodacre (2003) described a PCA model to differentiate pure extra virgin olive oil from adulterated olive oil with hazelnut oil, and a PLS model to quantify the amount of olive oil in the mixtures. Similar studies to the above mentioned ones are shown in Table 1. This table shows five papers using FTIR to detect adulteration of olive oil with other vegetable oil in blends binary, only Gurdeniz and Ozen (2009) develop a model to quantify olive oil in ternary blends. For Raman spectroscopy five works are reported, as in FTIR all the authors detect and quantify olive oil in blends binary, except Rohman and Che Man (2012) which quantifies olive oil in quaternary blends.

Table 1

The main disadvantage of the reported models to authenticate olive oil using spectroscopic techniques, such as FTIR and Raman, is the low number of different botanical species used to build the blends of olive oil with other edible vegetable oils. Most authors employ a small set of oils to elaborate the blends, and sometimes using a single olive oil or a limited number of vegetable edible oil (non-olive oil) in the different mixtures prepared. For example, Tay, Singh, Krishnan and Gore (2002) reported a method to authenticate olive oil using only thirty two olive oil and seven vegetable edible oils (non-olive oil) to build the different blends (Tay et al.,2002). Thus, the resulting models cannot be considered as global methods to detect adulteration of olive oil (independently of the cultivars) with any edible vegetable oil. Moreover, some authors erroneously apply PCA as discriminant analysis technique to develop and validate classification models of olive oil (Sun et al.,2015). PCA is an unsupervised data analysis technique used to explore the variability in the dataset and to evaluate if there are different groups of samples when the dimensionality of the data decreases. This exercise should not be used for classification purposes. In the literature there is only one published study where it is developed a classification model to distinguish

pure olive oil from other pure vegetable oil using FTIR or Raman spectroscopy. De la Mata et al. (2012) reported a partial least squares discriminant analysis (PLS-DA) aiming to distinguishing between olive oil and binary mixture of non-olive samples applying ATR-FTIR.

The aims of this study are: (i) discrimination of pure olive oil/non-pure olive oil, (ii) detection of adulterated olive oil and (iii) quantification of olive oil in blends (from binary to heptenary mixtures) with other vegetable edible oils using a number of chemometric techniques. For this purpose, we have developed a global and comprehensive analytical method to differentiate, detect and quantify olive oil in blends with any edible oils. The number of oils used in this work is wide, and spread worldwide. Although, in the "real world" the usual blends of olive oil with other seed oil are binary, a quality control laboratory does not know which was and/or how many were the seed oils used in adulteration, if any. For this reason, the proposed method aims at covering binary and higher-order blends which could be found.

2. MATERIALS AND METHODS

2.1. Chemicals

Isopropanol, *n*-hexane, methanol and tert-butyl methyl ether (TBME) were purchased from VWR International Eurolab, S.L. (Barcelona, Spain) and all of them were of HPLC grade. Other reagents, such as sodium methoxide, citric acid monohydrate, and anhydrous sodium sulphate were purchased from Merck (Darmstadt, Germany). The nitrogen (99.9999 %) used was provided by Air Liquid (Madrid, Spain).

2.2. Instrumentation

FT-IR spectra were obtained on a NICOLET iS5 spectrometer (Thermo Scientific, Waltham, Massachusetts, USA) equipped with a DTGS detector and KBr beam splitter. Spectra were obtained in the range of 4000 cm^{-1} to 550 cm^{-1} with a resolution of 2 cm^{-1} using a monolithic diamond attenuated total reflectance (ATR iD7) accessory. All the spectra were recorded at room temperature with 32 scans.

Raman measurements were carried out using IDRAMAN Reader (Ocean Optics, Oxford, UK) with 785 nm emission of a laser (23.4 mW at sample) for excitation. The laser was focused on the sample contained in 2 mL vial. For signal detection, a 2048-element NIR-enhanced CCD array with thermoelectric cooling to 10 °C was employed. An averaged spectrum for each sample was recorded in the range of 200 to 3200 cm^{-1} , using an integration time of 10 s each 3 scans.

NIR spectra were obtained using Antaris II (Thermo Electron Corporation, Waltham, Massachusetts, USA) FT-NIR analyzer, equipped with a diffuse reflection fibre optic and InGaAs detector. All the spectra, in the range of 4000 to 10000 cm^{-1} , were recorded at room temperature with 32 scans.

In all cases, each sample was analysed in triplicate.

2.3. Samples

Pure vegetable edible oils used to the classification models

67 samples of olive oils and 79 samples of other vegetable edible oils were used in this study. The samples of olive oils were constituted by 52 extra virgin olive oils (EVOO) samples, including 41 samples from 10 different monovarietals ("Arbequina", "Hojiblanca", "Picual", "Royal", "Manzanilla", "Cornicabra", "Empeltre", "Frantoio", "Verdial" and "Blanqueta") and 26 samples of varietal mixtures, 4 virgin olive oil samples (VOO), 5 olive oils, blend of virgin and refined (OO) and 6 pomace olive oil samples (POO). Vegetable edible oil samples (non-olive oils) consisted of 8 hazelnut

oils, 5 peanut oils, 10 canola oils, 2 safflower oils, 12 sunflower oils, 2 flax oils, 5 corn oils, 9 palm oils, 8 seeds oils (marketing mixture of unidentified seeds), 4 sesame oils, 8 soybean oils, 1 wheat oil and 4 grapeseed oils. In addition, a speciality olive oil extracted from previously dehydrated olive fruits was also added in this group. All samples were collected from marketed edible oils, purchased in food stores and sourced from respective partners from multiple geographical locations.

Blends of olive oil with other vegetable edible oils

To build the blends were used 27 olive oil samples, of which 22 EVOO (including 16 monovarietal oils), 3 VOO and 2 OO. In addition, 52 edible oils samples of 8 botanical origins, obtained each one from different suppliers, were used: 8 soybean oils, 11 sunflower oils, 10 rapeseed (canola) oils, 5 corn oils, 5 seeds oils (commercial blends of unknown seed oils), 5 peanut oils, 4 sesame oils and 4 grapeseed oils. Table 2 shows details on the composition of the different blends.

Table 2

All the oil samples were stored at 4 °C until the sample preparation in order to provide realistic testing conditions.

2.4. Sample preparation

Previous to the spectrometric analysis, a transesterification reaction was applied to the pure vegetable oil samples and all the different oil blends prepared. This reaction was carried out using 0.1 g/mL sodium methoxide in a methanol/TBME mixture, 4:6 (mL:mL), and then the extraction was performed with *n*-hexane. In this alkaline

medium, the free fatty acids presents in the oil are not methylated (Li & Watkins, 2001). A modification of the original procedure described by Biedermann *et al.* was applied (Bierdermann *et al.*, 1993). A detailed description of the procedure followed is described elsewhere (Jímenez-Carvelo, Pérez-Castaño, González-Casado & Cuadros-Rodríguez, 2017). The subsequent solution was stored at -25° C until analysis with less than 5% headspace under nitrogen.

2.5. Chemometrics

The FTIR and FT-NIR raw data files were exported to MATLAB (Mathworks, Massachusetts, USA, version R2013a). In order to reduce the variability associated to the intensity and derived from baseline, or other sources such as scattering effects, source or detector variations, or other general instrumental sensitivity effects, standard normal variate (SNV) and smoothing applying the Savitzky-Golay algorithm (second order polynomial filter with a 9-point window and first derivative) were used. Different chemometric tools have been applied for classification, including k-nearest neighbours (kNN), partial least squares discriminant analysis (PLS-DA), support vector machine-classification (SVM-C), one-class partial least squares (OCPLS) and soft independent modelling of class analogies (SIMCA). The classification results from each method have been evaluated on the basis of several quality metrics, such as: (i) sensitivity, (ii) specificity, (iii) positive (or precision) and negative predictive values, (iv) efficiency (or accuracy), (v) AUC (area under the receiver operating curve), (vi) Matthews correlation coefficient and (vii) Kappa coefficient. The meaning and way to calculate these metrics was recently reviewed [(Cuadros-Rodríguez, Pérez-Castaño & Ruiz-Samblás, 2016)].

Partial least squares regression (PLS-R) has been applied for quantification. Root Mean Square Error of Validation (RMSEV), Mean Absolute Error of Validation (MAEV) and Median Absolute Error of Validation (MdAEV) were used for accuracy assessment of the quantification methods.

198

199 *Olive/non-olive classification models*

200 Three classification strategies have been applied: (i) one input-class (1iC)
201 classification; (ii) two input-class (2iC) classification; and (iii) one input-class plus one
202 'dummy' class classification (or *pseudo* two input-class (*p*2iC) classification).

203 The main difference between the strategies pursued is the number of class used to
204 build the classification model. More detailed information can be found at the references
205 Jiménez-Carvelo, Pérez-Castaño, González-Casado and Cuadros-Rodríguez (2017)
206 and Jiménez-Carvelo, González-Casado, Pérez-Castaño and Cuadros-Rodríguez
207 (2017).

208 For each strategy applied (2iC, *p*2iC and 1iC) the original vector data set of pure
209 vegetable oil was divided into different groups to perform the classification model. The
210 selection was carried out using the Kennard-Stone (KS) algorithm [(Kennard & Stone,
211 1969)]. For 2iC, the training set was made up of 98 samples (44 olive oils and 54 non-
212 olive oils), and the remaining oil samples (23 olive oils and 25 non-olive oils) composed
213 the validation set. For *p*2iC, the training set which was made up of 61 samples (44 olive
214 oils and 17 analytical blanks), and the validation set composed by 102 samples (23
215 olive oils and 79 non-olive oils). For 1iC, the training set was composed by 44 olive oils
216 samples and the validation set by 102 samples (23 olive oils and 79 non-olive oils).

217 Once it was done, the classification models were developed. PLS_Toolbox (version
218 8.02, Eigenvector Research, Wenatchee, WA) for MATLAB environment was applied
219 for reducing of variables and classification methods: principal component analysis
220 (PCA) (Bro, 2014), k-nearest neighbours (kNN) (Steinbach & Tan, 2009), partial least
221 squares-discriminant analysis (PLS-DA) (Ballabio & Consonni, 2013) soft independent
222 modelling of class analogies (SIMCA) (Bylesjö et al., 2006), and support vector
223 machine-classification (SVM-C) (Luts et al., 2010). Moreover, one-class partial least

squares classification (OCPLS) (Xu, Yan, Cai & Yu, 2013) was performed applying the three variants of the function: (i) conventional ordinary linear OCPLS, (ii) nonlinear radial basis function (RBF) OCPLS, and (iii) partial robust M-regression (PMR) OCPLS, using the software provided by Xu *et al.*,(2013). All the options offered by the software were tested.

Adulterated olive oils detection models

Classification models to detect adulterations of olive oil with other vegetable edible oils were developed in order to apply a screening classification method previous to carry out the quantification. In this case, it was only applied the 2iC strategy. The training set was composed by 56 samples (44 pure olive oils and 12 adulterated olive oils) and validation set was made up of 35 samples (23 pure olive oils and 12 adulterated olive oils). As in the olive/non-olive classification models, PCA was used to reduce the variables, kNN, PLS-DA, SIMCA, SVM-C and OCPLS techniques were applied to developed the different classification models.

Olive oil quantification model

The original analytical data were divided in different groups to perform the statistical analysis. The calibration data set was made up of 18 samples whose adulteration levels were from 20 to 80 (g olive oil/100 g blend oil). The validation set for olive oil quantification was composed of 12 samples. The composition of the different samples is shown in Table 2. Partial least squares-regression (PLS-R) was used to build the model of quantitative prediction using standard parameters.

3. RESULTS AND DISCUSSION

Figure 1 shows a spectral fingerprint of the transesterified fraction of an EVOO sample with the three different spectroscopic techniques used. The FTIR and Raman fingerprints of EVOO show strong molecular vibrations and good variability between samples, the NIR fingerprints appear non selective. Therefore, classification and quantification models in FT-NIR were not developed due to the low specificity of the spectra of the transesterified fraction from the different vegetable edible oil samples.

Figure 1

3.1. Selection of variables

In order to reduce the number of variables and visualise the data a PCA model was obtained using FTIR and Raman fingerprints. In both spectroscopic techniques the selection of variable was performed examining the PCA loading plot. For that purpose, the regions of the spectra where the intensity of the loading was high were selected. Although the initial region of the Raman spectrum ($2900\text{-}2800\text{ cm}^{-1}$) shows a high value of the loadings, it was not finally selected since it did not improve the performance of the classification and quantification models.

The PCA model from FTIR data was developed with four principal components (PCs) which explain 98.87% of the variance. Figure 2 shows both the plot for FTIR spectrum and PCA loading plot with the three regions selected. The frequencies of the regions 1, 2 and 3 were $3100\text{-}2700\text{ cm}^{-1}$, $1800\text{-}1600\text{ cm}^{-1}$ and $1205\text{-}1080\text{ cm}^{-1}$ respectively.

Figure 2

The Raman spectra of all 146 samples were recorded. For the PCA model, four PCs were enough to explain 99.84% of the variance. Only the 950-650 cm^{-1} range, as shown in Figure 3, was chosen for analysis.

Figure 3

3.2. Olive/non-olive classification models

FTIR

In order to differentiate pure olive oils from other pure vegetable edible oils, different models were tested using the three regions selected; however, the best performance statistics were obtained for the models generated using the region 2.

The two-input class (2iC) strategy was used to develop the model applying the chemometric methods: kNN, PLS-DA, SVM-C and SIMCA. One-input class (1iC) strategy was applied when OCPLS and SIMCA models were performed and lastly, *pseudo* two-input class (*p*2iC) strategy was only applied to SIMCA model.

The target class was "olive oil" and the non-target class was "non-olive". In kNN, PLS-DA and SVM-C the olive class was assigned to samples with a predicted probability value equal to 1 and the non-olive class was defined by samples with a probability of 0. $K=3$ was enough to decide the neighbour distance in the kNN model. Classification of the samples of the validation set was performed directly by the software. There were only five samples misclassified, two olive oils samples and three non-olive oils samples (canola, peanut and hazelnut oils).

The PLS-DA model was built using six latent variables (LV), with 75.68% of the variance explained. Only one sample was not well classified corresponding to non-olive

oil (canola oil). The efficiency, area under the receiver operating curve (AUC) and Matthews's correlation coefficient were 0.98, 0.98 and 0.96 respectively. Figure 4 shows the classification plot obtained from the PLS-DA method.

Figure 4

The SVM-C model was developed optimizing the "C" and "nu" operational parameters. There are two commonly used versions of SVM classification, 'C-SVC' and 'nu-SVC'. "C" represents the penalty associated with errors. "Nu" is an alternative parameter for specifying the penalty associated with errors. It indicates a lower bound on the number of support vectors to use, given as a fraction of total calibration samples, and an upper bound on the fraction of training samples which are errors (misclassified) (SVM Function Settings, Eigenvector Documentation wiki. URL <http://wiki.eigenvector.com/index.php?title=Svmda>. Accessed 13.06.17). The results obtained in all the cases were similar. Moreover, all the models were tested with and without variable reduction using PCA and PLS. This variable reduction is named X-block compression by PLS_Toolbox software. The best results were obtained when an X-block compression with PLS was applied. The samples were directly classified by the software.

The 2iC, *p*2iC and 1iC strategies were tested to generate the SIMCA models. Classification of samples was performed using means of the normalised (also called, reduced) statistics values of residual- Q_r and Hotelling- T^2_r (Marini, 2010). The samples with values lower than 1 for both statistics were classified as olive oil. Firstly, a 2iC SIMCA classification was carried out. PCA model was built with 4 PCs and 5 PCs for olive and non-olive oil classes, respectively. Secondly, a *p*2iC SIMCA model was developed. In this case, 4 PCs and 3 PCs were chosen for olive oil and 'dummy'

classes respectively. At last, a 1iC SIMCA was performed with a 4 PCs for olive oil model.

An OCPLS class-model was also developed. Conventional OCPLS was built with 4 LVs, RBF OCPLS with 5 LVs and PMR OCPLS with 6 LVs. The regions for the samples classification were pre-established by the software.

To sum up, the 2iC strategy gave good results for all the discriminant methods. PLS-DA and SVM with reduction of variable using PLS were the best models; yielding the same classification results. The sensibility and specificity of all models were 1.00 and 0.96 respectively. In contrast, to SIMCA model lead to better classification results when 1iC strategy was used. The results for each model are shown in Table 3.

Table 3

Raman

In a similar way to FTIR, the 2iC strategy was applied with all the chemometric methods, *p*2iC strategy only with SIMCA and 1iC strategy with SIMCA and OCPLS. The classification criteria were the same as for FTIR with the different chemometric methods.

kNN classification model was built with k=3. Seven oil samples were misclassified (4 olive oils and 3 non-olive oils). The values of the quality performance metrics were similar with those obtained from FTIR models. Four LVs explaining 99.99% of the variance were enough to develop the PLS-DA model. This model was less efficient than PLS-DA model from FTIR.

As in the previous case of FT-IR spectra, SVM-C classification models were developed and tested with and without X-Block compression (reduction of variables). The results of all the models were the same excepting the (nu)-SVM-C model with reducing the

variable by PLS. In this model all the samples were classified in both classes (olive and non-olive oil classes) and the values of the quality performance metrics were not satisfactory.

The SIMCA classification model was built with 3 PCs and 4 PCs for olive and non-olive oil classes applying the 2iC strategy. On the contrary to FTIR model, this SIMCA model classified better the samples of the validation set. 3 PCs for each class were enough to develop the *p*2iC SIMCA model. This model classified all the oil samples in the class of non-olive oils. At last, the 1iC SIMCA model was built with 4 PCs for olive oil model.

OCPLS classification models were developed. In this case partial robust M-regression (PMR) OCPLS was the best model.

As in the case of FT-IR, the discriminant analysis methods gave good classification results; PLS-DA model was the best model. In contrast with the results for FTIR, SIMCA provided better results when the 2iC strategy was applied. Table 4 shows the results for each model.

Table 4

3.3. Adulterated olive oils detection models

Discriminant analysis and class-modelling methods were used for the discrimination of pure EVOO and EVOO adulterated with several vegetable edible oils. The chemometric techniques used and the criteria for classification were the same that to olive/non-olive classification models. Table 5 and 6 show the classification results of the different models tested from FTIR and Raman techniques. From FTIR, the best results were obtained when PLS-DA was applied. On the contrary, from Raman, the best models were obtained when SVM-C (optimizing with 'nu' operational parameter)

without and with X-Block compression by PLS was used. Only four EVOO adulterated samples were misclassified.

Table 5

Table 6

3.4. Olive oil quantification model

Quantitative analysis of blends of olive oil with other vegetable edible oils was performed building a specific PLS-R model from FTIR and Raman fingerprints on the regions previously selected (see section 3.1). In order to achieve more realistic conditions of the composition of olive oil, the proportion of olive oil in the blends of the training and validation set is different, in contrast to some research work about quantification of olive oil using spectroscopic techniques in which the composition is similar in both set.

The reliability of the different models was established on the basis of: (i) the determination coefficient (R^2) and (ii) the errors of quantification (validation errors) were evaluated with the Root Mean Square Error of Validation (RMSEV), Mean Absolute Error of Validation (MAEV) and Median Absolute Error of Validation (MdAEV) (Hyndman & Koehler, 2006; ASTM E1655-05, 2012). The results obtained (g EVOO/100 g blend) in terms of R^2 , RMSEV, MAEV and MdAEV were 0.86, 17.6, 14.6 and 16.0 respectively from FTIR and 0.93, 34.2, 27.8 and 29.6 respectively from Raman. Figure 5 shows the concentration values obtained from the PLS model vs. the actual concentration of any vegetable edible oil in olive oil samples using FTIR-ATR.

Figure 5

Although the R^2 obtained from FTIR is not sufficiently good, the validation errors (about 15-17%) are better than the validation errors obtained from Raman (about 28-34%).

4. CONCLUSION

Methyl-transesterified provides the information needed to authenticate of olive oil. The method developed could be named “global method” of detection, discrimination and quantification of olive oil in blends with other vegetable edible oils. Moreover, due to a transesterification step prior to spectroscopic analysis the problem of the low selectivity of these techniques has been resolved. Using FTIR and applying PLS-DA is performed without the need of any resource intensive chromatographic analysis. Discriminant analysis classified well the 100% olive oils samples and in addition, the proportion of olive oil in blends with other vegetable edible oils has been successively quantified using PLS-R.

407

ABBREVIATIONS AND ACRONYMS

408

1iC, one input-class classification

2iC, two input-class classification

ATR, attenuated total reflectance

AUC, area under the receiver operating curve

Bay-LS-SVM, Bayesian-least squares-support vector machine

BOM, bean with omega

CAM, camellia oil

CAN, canola oil

COG, corn germ oil

COR, corn oil

COT, cottonseed oil

EVOO, extra virgin olive oil

FLA, flaxseed oil

FT-IR, Fourier transform-infrared spectroscopy

FT-NIR, Fourier transform-near infrared spectroscopy

GAR, garlic oil

GSO, Grapeseed oil

HAZ, hazelnut oil

kNN, k-nearest neighbors

KS, Kennard-Stone

LDA, linear discriminant analysis

LS-SVM, least squares-support vector machine

LV, latent variables

MAE, Mean absolute error

MAEV, mean absolute error of validation

MdAEV, median absolute error of validation

MDVO, mixture of different vegetable edible oils (non-olive oil)

MLR, multiple linear regression

OCPLS, one class partial least squares classification

OO, olive oil

$p2iC$, *pseudo* two input-class classification

PC, principal component

PCA, principal component analysis

PCR, principal component regression

PLS-DA, partial least squares-discriminant analysis

PLS-R, partial least squares regression

PLS-R, partial least squares regression

PMR, partial robust M-regression

POO, pomace olive oil

R^2 , determination coefficient

RBF, radial basis function

RBO, rice bran oil

RMSE, Root mean square error

RMSEV, root mean square error of validation

RPS, rapeseed oil

SAF, safflower oil

SES, sesame oil

SIMCA, soft independent modelling of class analogy

SNV, standard normal variate

SOY, soybean oil

SUN, sunflower oil

SVM-C, support vector machine classification

TBME, tert-butyl methyl ether

VOO, virgin olive oil

WGE, wheat germ oil

WO, walnut oil

409

410

411 REFERENCES

412

- Regulation (UE) N° 29/2012 on marketing standards for olive oil (consolidated version 2012R0029), Official Journal of the European Union. (2016) 005.001, 1-15
- ASTM E1655-05(2012). (2012). Standard practices for infrared multivariate quantitative analysis. ASTM International, West Conshohocken (PA).
- Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Method*, 5, 3790–3798.
- Biedermann, M., Grob, K., & Mariani, C. (1993). Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. *European Journal of Lipid Science and Technology*, 95, 127–133.
- Bro, R., & Smilde, A. (2014). Principal component analysis. *Analytical Method*, 6, 2812–2831.
- Bytesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics*, 20, 341–351.
- CODEX STAN 33-1981. Revision 2015. Standard for olive oil oils and olive pomace oils. Codex Alimentarius, FAO/WHO
- Commission Regulation (EEC) N° 2568/1991 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis (consolidated version 01991R2568), Official Journal of the European Union (2016), 031.001, 1-127
- Cuadros-Rodríguez, L., Pérez-Castaño, E., & Ruiz-Samblás, C. (2016). Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends in Analytical Chemistry*, 80, 612-624.
- De la Mata, P., Dominguez-Vidal, A., Bosque-Sendra, J.M., Ruiz-Medina, A., Cuadros-Rodríguez, L., & Ayora-Cañada, M. (2012). Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics. *Food Control*, 23, 449–455.
- Dong, W., Zhang, Y., Zhang, B., & Wang X. (2012). Quantitative analysis of adulteration of extra virgin olive oil using Raman spectroscopy improved by Bayesian framework least squares support vector machine. *Analytical Method*, 4, 2772–2777.
- El-Abassy, R.M., Donfack, P., & Materny A. (2009). Visible Raman spectroscopy for the discrimination of olive oils from different vegetable oils and the detection of

- adulteration. *Journal of Raman Spectroscopy*, 40, 1284–1289
- Gurdeniz, G., & Ozen B. (2009). Detection of adulteration of extra virgin olive oil by chemometric analysis of mid-infrared spectral data. *Analytical Method*, 116, 519–525
- Hyndman, R.J., & Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688.
- Jiménez-Carvelo, A.M., González-Casado, A., Pérez-Castaño, E., & Cuadros-Rodríguez L. (2017). Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods. *Journal of AOAC International*, 100(2), 1-6
- Jiménez-Carvelo, A.M., Pérez-Castaño, E., González-Casado, A., & Cuadros-Rodríguez, L. (2017). One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction. *Food Chemistry*, 221, 1784–1791
- Kennard, R.W., & Stone, L.A. (1969). Computer aided design of experiments. *Technometrics*, 11, 137–148.
- Lerma-García, M.J., Ramis-Ramos, G., Herrero-Martínez, J.M., & Simó-Alfonso, E.F. (2010). Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. *Food Chemistry*, 118, 78–83.
- Li, Y. & Watkins, B.A. (2001). Unit 1.2: Analysis of Fatty Acids in Food Lipids, in Whitaker, J. (Ed.), *Current Protocols in Food Analytical Chemistry*. John Wiley and Sons, Inc., New York, D1.2.1-D1.2.15.
- López-Díez, E.C., Bianchi, G., & Goodacre, R. (2003). Rapid and quantitative assessment of the adulteration of virgin olive oils with hazelnut oils using Raman spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry*, 51, 6145–6150.
- Luts, J., Ojeda, F., De Plas, R.V., De Moor, B., Huffel, S.V., & Suykens, J.A.K. (2010). A tutorial on support vector machine-based methods for classifications problems in chemometrics. *Analytical Chimica Acta*, 665, 129–145.
- Marini, F. (2010). Classification methods in chemometrics. *Current Analytical Chemistry*, 6, 72-79.
- Regulation (UE) N° 1169/2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives

2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004, Official Journal of the European Union (2011), L 304/18

Rohman, A., Che, & Man, Y.B. (2011). Determination of extra virgin olive oil in quaternary mixture using FTIR spectroscopy and multivariate calibration. *Spectroscopy*, 26, 203–211.

Rohman, A., & Che Man, Y.B. (2012). Authentication of extra virgin olive oil from sesame oil using FTIR spectroscopy and gas chromatography. *International Journal of Food Properties*, 15, 1309–1318.

Steinbach, M., & Tan, P.N. (2009). kNN: k-Nearest neighbors. In: Wu X., Kumar V, (Eds), *The top ten algorithms in data mining* (pp. 151–161). Boca Raton (FL): Chapman & Hall/CRC.

Sun, X., Lin, W., Li, X., Shen, Q., & Luo, H. (2015). Detection and quantification of extra virgin olive oil adulteration with edible oils by FT-IR spectroscopy and chemometrics. *Analytical Methods*, 7, 3939–3945.

SVM Function Settings, Eigenvector Documentation Wiki [accessed 13 June 2017], <http://wiki.eigenvector.com/index.php?title=Svmlda>.

Tay, A., Singh, R.K., Krishnan, S.S., & Gore, J.P. (2002). Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared spectroscopy. *LWT – Food Science and Technology*, 35, 99-103.

Tiryaki, G.Y., & Ayvaz, H. (2017). Quantification of soybean oil adulteration in extra virgin olive oil using portable Raman spectroscopy. *Journal of Food Measurement and Characterization*, 11, 523-529.

Xu, L., Yan, S.M., Cai, C.B., & Yu, X.P. (2013). One-class partial least squares (OCPLS) classifier. *Chemometrics and Intelligent Laboratory System*, 126, 1–5.

Zou, M.Q., Zhang, X.F., Qi, X.H., Ma, H.L., Dong, Y., Liu, C.W., Guo, X., & Wang H. (2009). Rapid authentication of olive oil adulteration by Raman spectrometry. *Journal of Agricultural and Food Chemistry*, 57, 6001–6006.

Table 1. Chemometric methods using FTIR or Raman for the authentication of olive oil found in the literature

| Nº | Analytical technique | Amount and types of edible oils | Blends | Aims | Chemo-metrics | Results (Quality Features) | Ref. |
|----|--|--|---|---|---------------|---|--|
| 1 | FTIR (4000-650 cm ⁻¹) | EVOO ^t (40), CAM ^a (5), SOY ^b (5), SUN ^c (5) and COR ^d (5) oils | EVOO-CAM, EVOO- SOY, EVOO-SUN, EVOO-COR | Classification model of EVOO and binary blends of EVOO with edible oil Quantification of EVOO in binary blends | PCA and PLS | R ² : 0.98 - 0.99 RMSE: 1.9 % (EVOO-SUN); 9.5% (EVOO-CA); 1.72 % (EVOO-SOY); 2.2% (EVOO-COR) | {{Sun,Lin, Li, Shen & Luo, 2015}} |
| 2 | FTIR (1200-900 and 2949-2885 cm ⁻¹) | EVOO (1), GSO ^e (1), RBO ^f (1), WO ^g (1) oils | GSO-WO, EVOO-RBO, EVOO-RBO-GSO-WO | Quantification of EVOO in quaternary mixture | PLS | R ² : 0.99 ; RMSE : 3.7% | {{Rohman & Che Man, 2011}} |
| 3 | FTIR (1207-1018, 1517-1222 and 3050-2927 cm ⁻¹) GC | EVOO, CAN ^h , COR, GSO, SOY, SES ⁱ , SUN and WO oils | EVOO-SES | Classification model of EVOO and other pure edible oil based on their fatty acids profiles. Quantification of EVOO in blends of EVOO-SES | PLS and PCR | R ² : 1.00 ; RMSE : 7.0% (PLS) R ² : 0.997; RMSE: 1.1% (PCR) | {{Rohman & Che Man, 2012}} |
| 4 | FTIR (4000-1000 cm ⁻¹) | EVOO (6), HAZ ^j (6), SUN (6), COR (3), COG ^k (2) and SOY (6) oils | EVOO-HAZ, EVOO-SUN, EVOO-CORN, EVOO-SOY | Classification of vegetable oils using LDA Determination of EVOO adulteration | LDA and MLR | R ² : 0.91; MAE: 2.0 (EVOO-HAZ) R ² : 0.99; MAE: 1.7 (EVOO-SUN) R ² : 0.99; MAE: 1.5 (EVOO-CORN) R ² : 0.98; MAE: 1.9 (EVOO-SOY) | {{Lerma-García, Ramis-Ramos, Herrero-Martínez & Simó-Alfonso, 2010}} |

| | | | | | | | |
|----|---|---|--|---|----------------------------|--|---|
| 5 | FTIR (3080-2800 cm ⁻¹) | EVOO (25), COR, SUN, RPS and COT ^s oils | EVOO-SUN-COR, EVOO-COT, EVOO-RPS | Classification model of EVOO and adulterated EVOO Quantification of EVOO in the mixtures | SIMCA and PLS | R ² : 0.99; RMSE: 10.4% (EVOO-SUN-COR) R ² : 0.95; RMSE: 14.0% (EVOO-COT) R ² : 0.93; RMSE: 13.2% (EVOO-RPS) | {{Gurdeniz & Ozen, 2009}} |
| 6 | Raman (1000-3000 cm ⁻¹) | EVOO (31) and HAZ (10) oils | EVOO-HAZ | Quantification of EVOO in blends with hazelnut oil | PCA and PLS | R ² = 0.98 RMSE 10.94% | {{López-Diez , Bianchi & Goodacre, 2003}} |
| 7 | Raman (800-1800 and 2850-3020 cm ⁻¹) | EVOO (18), RPS ^l , SES, GAR ^m , BOM ⁿ , SUN (3), WO, SAF ^o (2), SOY, WGE ^p , and FLA ^r oils | EVOO-SUN#1, EVOO-SUN#2, EVOO-SUN#3 | Discrimination model of EVOO and adulterated EVOO. Estimation of the SUN oil content in EVOO | PCA and PLS | R ² : 0.99; RMSE(cross-valid): 9.81% R ² : 0.99; RMSE(cross-valid): 9.88% R ² : 0.98; RMSE (cross-valid): 9.71% | {{El-Abassy, Donfack & Materny, 2009}} |
| 8 | Raman (1000-1800 cm ⁻¹) | EVOO (6), POO ^u (1), SOY (3), SUN (3), RPS (2) and COR (2) oils | EVOO-SOY, EVOO-SUN, EVOO-RPS, EVOO-COR | Model to detect adulterated EVOO | PCA | Intensity ratio | {{Zhou et al., 2009}} |
| 9 | Raman (800-1800 cm ⁻¹) | EVOO (5), SOY (3), corn (3) and SUN (3) oils | EVOO-SUN, EVOO-SOY, EVOO-COR | Quantification of EVOO in binary blends | Bay-LS-SVM, LS-SVM and PLS | R ² : 0.99; RMSE: 5.1% (Bay-LS-SVM) R ² : 0.99; RMSE: 6.9% (LS-SVM) R ² : 0.99; RMSE: 8.4% (PLS) | {{Dong, Zhang, Zhang & Wang, 2012}} |
| 10 | Raman (1000-1800 cm ⁻¹) | EVOO and SOY oil | EVOO-SOY | Quantification of SOY adulteration in EVOO | PLS | R ² = 0.99; RMSE: 1.3% | {{Tiryaki & Ayvaz, 2016}} |

414 ^a Camellia; ^b Soybean; ^c Sunflower; ^d Corn; ^e Grapeseed; ^f Rice bran; ^g Walnut; ^h Canola; ⁱ Sesame; ^j Hazelnut; ^k Corn germ; ^l Rape seed; ^m Garlic; ⁿ Bean with Omega 3; ^o Safflower; ^p Wheat germ; ^r
415 flaxseed; ^scottonseed, ^tExtra virgin olive oil, ^u Pomace olive oil

416 RMSE: Root mean square error; MAE: Mean absolute error

Table 2. Percentage and composition of the olive oil and other vegetable edible oil in the oil blend samples.

| N° | Composition |
|---|--|
| <i>(a) Calibration set</i> | |
| 1 100% MDVO ^a | 20% seed#1 oil, 20% peanut#1 oil, 20% sunflower#1 oil, 20% canola#1 oil and 20% corn#1 oil |
| 2 100% MDVO | 20% soybean#1 oil, 20% soybean#2 oil, 20% sunflower#2 oil, 20% canola#2 oil and 20% grapeseed#1 oil |
| 3 100% MDVO | 20% seed#2 oil, 20% sesame#1 oil, 20% peanut#2 oil, 20% corn#2 oil and 20% grapeseed#2 oil |
| 4 20% EVOO ^b + 80% MDVO | 10% EVOO#1, 5% EVOO#2, 5% EVOO#3, 13% soybean#3 oil, 13% canola#3 oil, 13% corn#1 oil, 13% seed#3 oil, 13% grapeseed#3 oil and 13% peanut#2 oil |
| 5 15% EVOO + 5% OO ^c + 80% MDVO | 5% OO#4, 5% EVOO#5, 5% EVOO#6, 5% EVOO#7, 8% sunflower#3 oil, 8% sunflower#4 oil, 8% canola#1 oil, 8% canola#4 oil, 16% corn#3 oil, 16% sesame#2 oil and 16% peanut#3 oil |
| 6 15% EVOO + 5% VOO ^d + 80% MDVO | 5% EVOO#8, 5% EVOO#9, 5% EVOO#11, 5% VOO#10, 13% sunflower#5 oil, 13% sunflower#2, 26% corn#2 oil and 26% grapeseed#2 oil |
| 7 30% EVOO + 10% OO + 60% MDVO | 10% OO#12, 10% EVOO#5, 10% EVOO#7, 10% EVOO#14, 15% soybean#4 oil, 15% canola#5 oil, 15% seed#3 oil and 15% peanut#4 oil |
| 8 30% EVOO + 10% OO + 60% MDVO | 10% EVOO#6, 10% EVOO#13, 10% EVOO#8, 10% OO#4, 15% sunflower#6 oil, 15% canola#5 oil, 15% corn#4 oil and 15% grapeseed#2 oil |
| 9 30% EVOO + 10% VOO + 60% MDVO | 10% EVOO#1, 10% EVOO#3, 10% EVOO#2, 10% VOO#15, 15% sunflower#7 oil, 15% corn#1 oil, 15% sesame#1 oil and 15% peanut#4 oil |
| 10 60% EVOO + 40% MDVO | 15% EVOO#6, 15% EVOO#7, 15% EVOO#13, 15% EVOO#14, 5% soybean#1 oil, 5% soybean#5 oil, 10% canola#6 oil, 10% sesame#3 oil and 10% grapeseed#4 oil |
| 11 36% EVOO + 12% OO + 12% VOO + 40% MDVO | 12% EVOO#6, 12% EVOO#2, 12% EVOO#5, 12% OO#12, 12% VOO#15, 8% canola#7 oil, 8% corn#5 oil, 8% seed#4 oil, 8% grapeseed#8 oil and 8% peanut#3 oil |
| 12 40% EVOO + 10% VOO + 10% OO + 40% MDVO | 10% EVOO#9, 10% EVOO#11, 10% EVOO#1, 10% EVOO#8, 10% VOO#10, 10% OO#12, 7% sunflower#8 oil, 6.6% canola#8 oil, 6.6% corn#2 oil, 6.6% sesame#2 oil, 6.6% seed#2 oil and 6.6% peanut#5 oil |
| 13 40% EVOO + 20% VOO + 20% OO + 20% MDVO | 20% EVOO#5, 20% EVOO#2, 20% VOO#15, 20% OO#12, 5% sunflower#9 oil, 5% corn#3 oil, 5% seed#1 oil and 5% grapeseed#3 oil |
| 14 80% EVOO + 20% MDVO | 30% EVOO#6, 25% EVOO#7, 25% EVOO#9, 5% seed#5 oil, 5% peanut#1 oil, 5% canola#9 oil and 5% canola#2 oil |
| 15 60% EVOO + 20% OO + 20% MDVO | 20% EVOO#11, 20% EVOO#1, 20% EVOO#13, 20% OO#12, 5% soybean#6 oil, 5% corn#1 oil, 5% sesame#4 oil and 5% grapeseed#1 oil |

| | | |
|----|-----------|--------------|
| 16 | 100% EVOO | 100% EVOO#16 |
| 17 | 100% EVOO | 100% EVOO#17 |
| 18 | 100% VOO | 100% VOO#18 |

(b) Validation set

| | | |
|----|-----------------------------|---|
| 1 | 68% EVOO + 32% MDVO* | 68% EVOO#6, 25% corn#5 oil, 3% peanut#3 oil and 4% grapeseed#4 oil |
| 2 | 17.50% VOO + 82.50% MDVO | 17.50% VOO#15, 17% sunflower#8, 11% soybean#4 oil, 28% canola#6 oil, 26% peanut#1 oil and 0.5% seed#5 oil |
| 3 | 93% VOO + 7% MDVO | 93% VOO#19, 2% corn#3 oil and 5% sesame#4 oil |
| 4 | 44% EVOO + 56% MDVO | 44% EVOO#20, 13% peanut#3 oil, 8% canola#5 oil and 35% canola#4 oil |
| 5 | 5% EVOO + 95% MDVO | 5% EVOO#27, 40% canola#9 oil, 23% soybean#2 oil, 7% grapeseed#2 oil, 15% canola#4 oil and 10% sunflower#3 oil |
| 6 | 68% EVOO + 32% MDVO | 68% EVOO#21, 10% sesame#4 oil, 7% soybean#7 oil and 15% seed#3 oil |
| 7 | 70% VOO + 30% MDVO | 70% VOO#10, 1% sunflower#2 oil, 9% sesame#1 oil, 17% corn#1 oil and 3% sunflower#1 |
| 8 | 31% EVOO + 69 MDVO | 31% EVOO#22, 24% sunflower#3 oil, 13% sesame#4 oil, 20% soybean#7 oil, 2% peanut#5 oil and 10% grapeseed#2 |
| 9 | 52% EVOO + 48% MDVO | 52% EVOO#23, 28% canola#7, 13% soybean#6 oil, 5% grapeseed#1 oil and 2% sesame#4 oil |
| 10 | 25% EVOO + 75% MDVO | 25% EVOO#24, 25% corn#1 oil, 25% sunflower#2 oil and 25% peanut#3 oil |
| 11 | 90% EVOO + 10% MDVO | 90% EVOO#25, 5% canola#2 oil and 5% soybean#2 oil |
| 12 | 40% EVOO + 60% MDVO | 40% EVOO#26, 30% peanut#1 oil and 30% canola#6 |

417 ^a MDVO: Mixture of different vegetable edible oils (non-olive oils)

418 ^b EVOO: Extra virgin olive oil

419 ^c OO: Olive oil

420 ^d VOO: Virgin olive oil

421

Table 3. Values of the quality performance metrics from the different FTIR olive/non-olive classification methods

| Performance features Region 2 (1680-1800 cm ⁻¹) | kNN | PLS-DA | (nu)SVM-C | | | (c)SVM-C | | | SIMCA | | | OCPLS | | |
|--|------|--------|-----------|------|------|----------|------|------|-------|------|------|---------------------|------|------|
| | | | None | | | None | | | | | | Ordinar y linear | RBF | PRM |
| | | | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | p2iC | 1iC | 1iC | 1iC | 1iC |
| Sensibility (Recall) | 0.91 | 1.00 | 1.00 | 0.83 | 1.00 | 1.00 | 0.87 | 1.00 | 0.91 | 0.87 | 0.96 | 0.74 | 0.87 | 0.57 |
| Specificity | 0.88 | 0.96 | 0.88 | 0.88 | 0.96 | 0.92 | 0.84 | 0.96 | 0.32 | 0.76 | 0.71 | 0.80 | 0.20 | 0.94 |
| Positive predictive value (Precision) | 0.88 | 0.96 | 0.88 | 0.86 | 0.96 | 0.92 | 0.83 | 0.96 | 0.55 | 0.51 | 0.49 | 0.52 | 0.24 | 0.72 |
| Negative predictive value | 0.92 | 1.00 | 1.00 | 0.85 | 1.00 | 1.00 | 0.88 | 1.00 | 0.80 | 0.95 | 0.98 | 0.91 | 0.84 | 0.88 |
| Efficiency (Accuracy) | 0.90 | 0.98 | 0.94 | 0.85 | 0.98 | 0.96 | 0.85 | 0.98 | 0.60 | 0.78 | 0.76 | 0.78 | 0.35 | 0.85 |
| AUC (Correctly classified rate) | 0.90 | 0.98 | 0.94 | 0.85 | 0.98 | 0.96 | 0.85 | 0.98 | 0.62 | 0.81 | 0.83 | 0.77 | 0.54 | 0.75 |
| Matthews correlation coefficient | 0.79 | 0.96 | 0.88 | 0.71 | 0.96 | 0.92 | 0.71 | 0.96 | 0.29 | 0.54 | 0.56 | 0.48 | 0.08 | 0.55 |
| Kappa coefficient | 0.79 | 0.96 | 0.88 | 0.71 | 0.96 | 0.92 | 0.71 | 0.96 | 0.23 | 0.50 | 0.50 | 0.46 | 0.04 | 0.54 |

422

423

Table 4. Values of the quality performance metrics from the different Raman olive/non-olive classification methods

| Performance features (650-950 cm ⁻¹) | kNN | PLS-DA | (nu)SVM-C | | | (c)SVM-C | | | SIMCA | | | OCPLS | | |
|---|------|--------|-----------|------|------|----------|------|------|-------|------|------|----------|-------|------|
| | | | None | PCA | PLS | None | PCA | PLS | 2iC | p2iC | 1iC | Ordinar | RBF | PRM |
| | | | | | | | | | | | | y linear | | |
| | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | | | | 1iC | 1iC | 1iC |
| Sensibility (Recall) | 0.83 | 0.88 | 0.83 | 0.83 | 1.00 | 0.83 | 0.83 | 0.83 | 0.67 | 0.00 | 0.50 | 0.42 | 0.46 | 0.33 |
| Specificity | 0.88 | 0.88 | 0.88 | 0.88 | 0.00 | 0.88 | 0.88 | 0.88 | 0.88 | 1.00 | 0.94 | 0.29 | 0.23 | 0.95 |
| Positive predictive value (Precision) | 0.87 | 0.88 | 0.87 | 0.87 | 0.49 | 0.87 | 0.87 | 0.87 | 0.84 | - | 0.71 | 0.15 | 0.15 | 0.67 |
| Negative predictive value | 0.85 | 0.88 | 0.85 | 0.85 | - | 0.85 | 0.85 | 0.85 | 0.73 | 0.76 | 0.86 | 0.62 | 0.58 | 0.82 |
| Efficiency (Accuracy) | 0.86 | 0.88 | 0.86 | 0.86 | 0.49 | 0.86 | 0.86 | 0.86 | 0.78 | 0.76 | 0.83 | 0.32 | 0.28 | 0.80 |
| AUC (Correctly classified rate) | 0.86 | 0.88 | 0.86 | 0.86 | 0.50 | 0.86 | 0.86 | 0.86 | 0.77 | 0.50 | 0.72 | 0.36 | 0.34 | 0.64 |
| Matthews correlation coefficient | 0.71 | 0.76 | 0.71 | 0.71 | - | 0.71 | 0.71 | 0.71 | 0.56 | - | 0.50 | -0.25 | -0.29 | 0.37 |
| Kappa coefficient | 0.71 | 0.76 | 0.71 | 0.71 | 0.00 | 0.71 | 0.71 | 0.71 | 0.55 | 0.00 | 0.48 | 0.24 | -0.19 | 0.34 |

The hyphen "-" is signifying that the performance feature cannot be determined

Table 5 Values of the quality performance metrics from the different FTIR adulterated olive oils detection models

| Performance features | kNN | PLS-DA | SIMCA | (nu)SVM-C | | | (c)SVM-C | | |
|---|-------|--------|-------|-----------|------|------|----------|------|------|
| | | | | None | PCA | PLS | None | PCA | PLS |
| | | | | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC |
| Region 2 (1680-1800 cm⁻¹) | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC |
| Sensibility (Recall) | 0.83 | 0.70 | 0.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.87 |
| Specificity | 0.50 | 0.92 | 0.58 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.67 |
| Positive predictive value (Precision) | 0.76 | 0.94 | 0.76 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.83 |
| Negative predictive value | 0.60 | 0.61 | 0.50 | - | - | - | - | - | 0.73 |
| Efficiency (Accuracy) | 0.71 | 0.77 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 | 0.80 |
| AUC (Correctly classified rate) | 0.66 | 0.81 | 0.64 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.77 |
| Matthews correlation coefficient | 0.34 | 0.58 | 0.27 | - | - | - | - | - | 0.55 |
| Kappa coefficient | -0.11 | 0.55 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.55 |

The hyphen "-" is signifying that the performance feature cannot be determined

Table 6. Values of the quality performance metrics from the different Raman adulterated olive oils detection models

| Performance features (650-950 cm ⁻¹) | kNN | PLS-DA | SIMCA | (nu)SVM-C | | | (c)SVM-C | | |
|---|-------|--------|-------|-----------|------|------|----------|-------|------|
| | | | | None | PCA | PLS | None | PCA | PLS |
| | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC | 2iC |
| Sensibility (Recall) | 0.83 | 0.92 | 0.46 | 1.00 | 1.00 | 1.00 | 0.88 | 0.71 | 0.83 |
| Specificity | 0.08 | 0.67 | 0.33 | 0.67 | 0.00 | 0.67 | 0.67 | 0.00 | 0.67 |
| Positive predictive value (Precision) | 0.65 | 0.85 | 0.58 | 0.86 | 0.67 | 0.86 | 0.84 | 0.59 | 0.83 |
| Negative predictive value | 0.20 | 0.80 | 0.24 | 1.00 | - | 1.00 | 0.73 | 0.00 | 0.67 |
| Efficiency (Accuracy) | 0.58 | 0.83 | 0.42 | 0.89 | 0.67 | 0.89 | 0.81 | 0.47 | 0.78 |
| AUC (Correctly classified rate) | 0.46 | 0.79 | 0.40 | 0.83 | 0.50 | 0.83 | 0.77 | 0.35 | 0.75 |
| Matthews correlation coefficient | -0.11 | 0.61 | -0.20 | 0.76 | - | 0.76 | 0.55 | -0.35 | 0.50 |
| Kappa coefficient | -0.10 | 0.61 | -0.19 | 0.73 | 0.00 | 0.73 | 0.55 | -0.33 | 0.50 |

The hyphen "-" is signifying that the performance feature cannot be determined

FIGURE CAPTIONS

Figure 1. Examples of vibrational spectra of extra virgin olive oil (EVOO) acquired from:
(a) FTIR (b) Raman and (c) FT-NIR

Figure 2. (a) Superposed FTIR spectra and (b) loading plot of the 146 vegetable oil samples showing the three regions selected.

Figure 3. Plot of (a) superposed Raman spectra and (b) loading plot of all the vegetable edible oil samples showing the region selected.

Figure 4. Classification plot from FTIR applying two input-class (2iC) classification strategy on PLS-DA. The green squares (■) and the red triangles (▼) represent the olive and non-olive class, respectively.

Figure 5. Concentration values for adulteration obtained from the PLS model vs. the actual concentration of olive oil using FTIR-ATR.

Figure 1

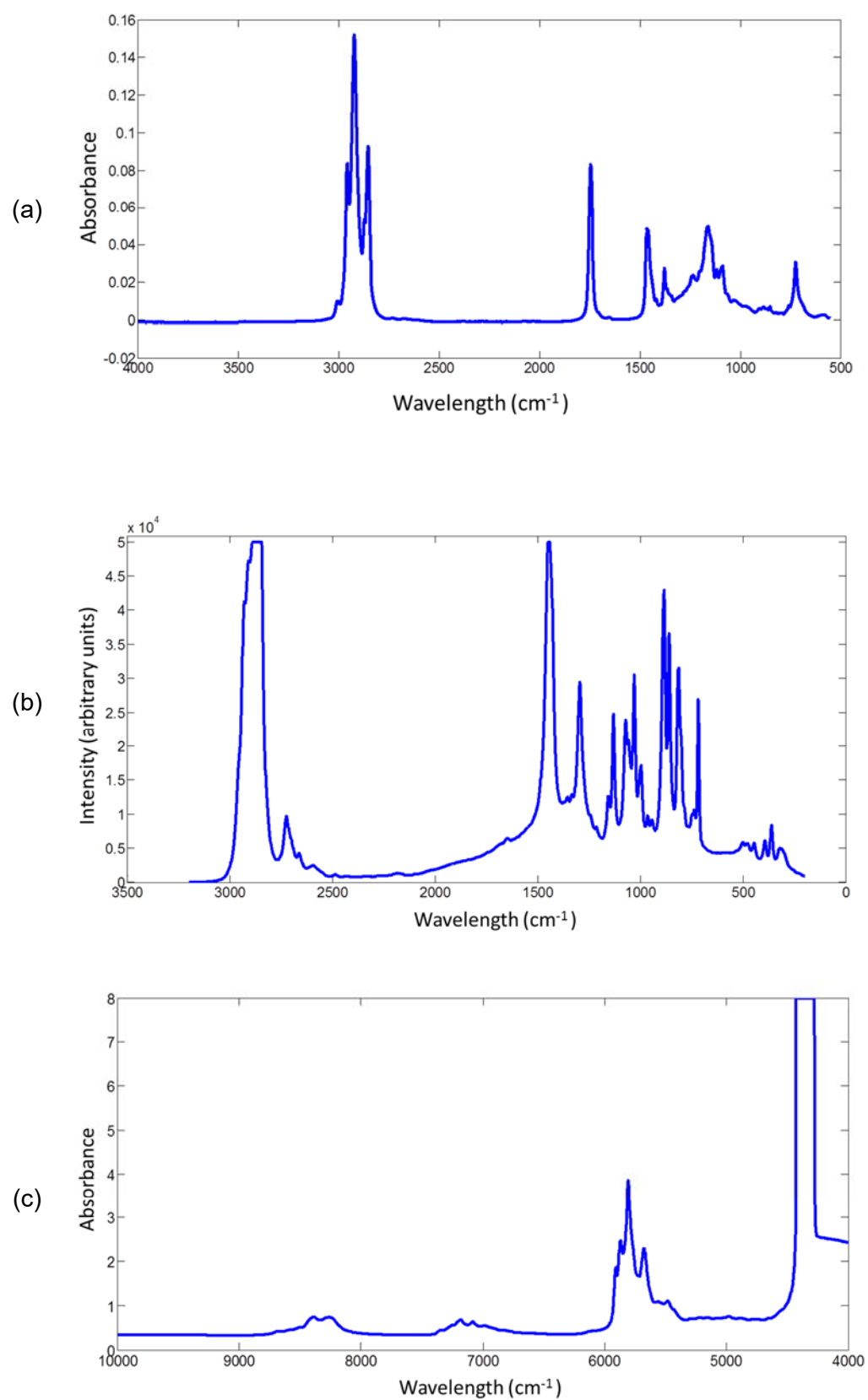
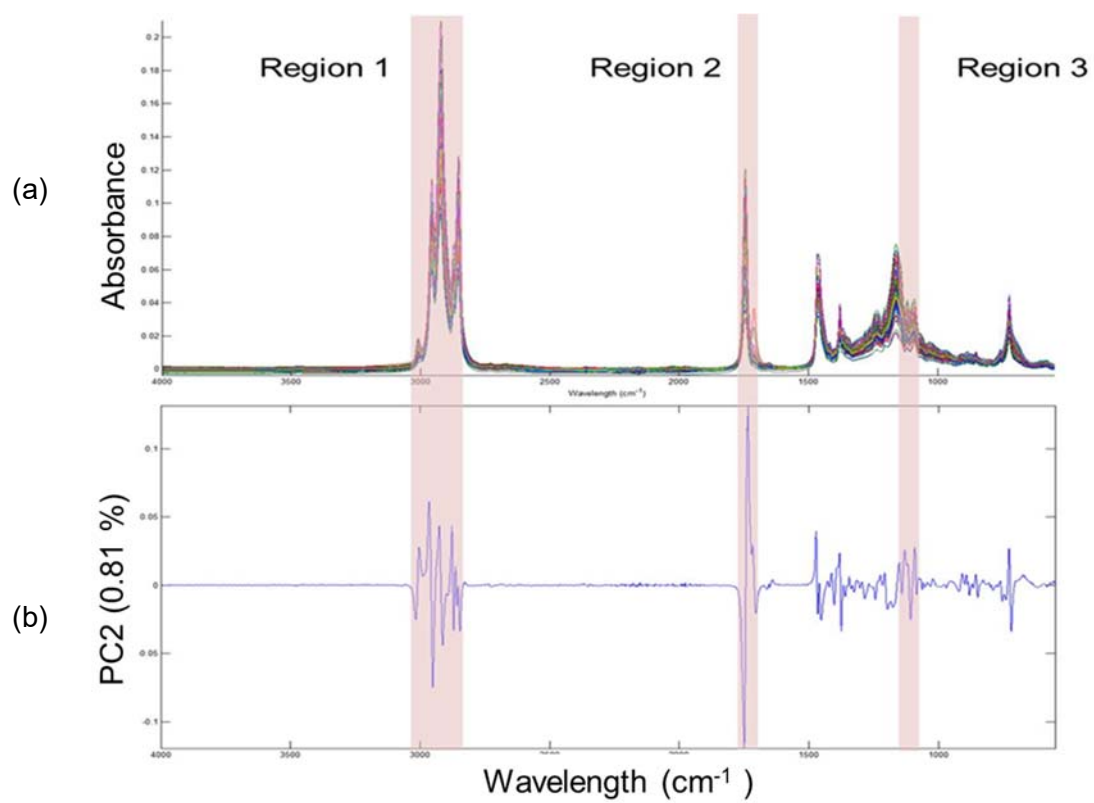
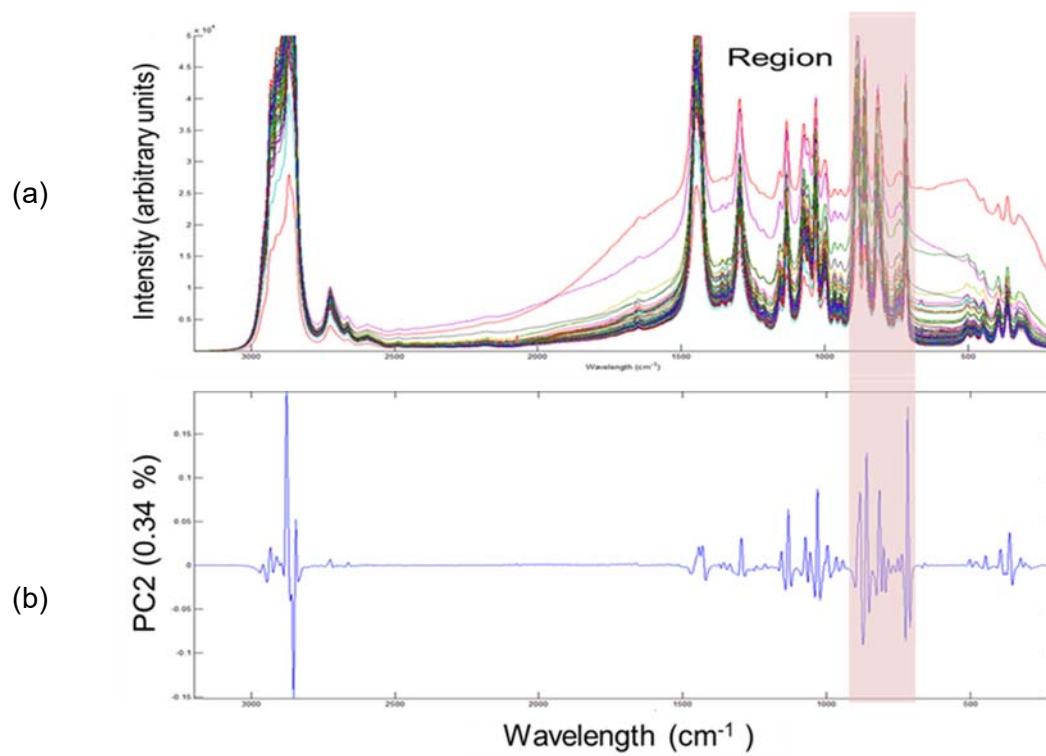


Figure 2



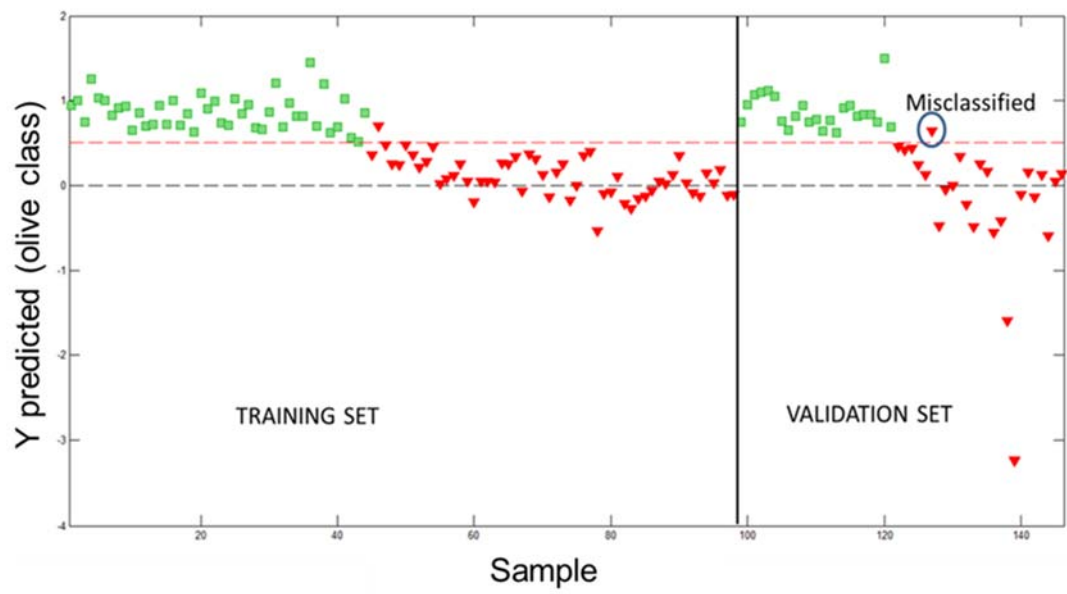
448

Figure 3



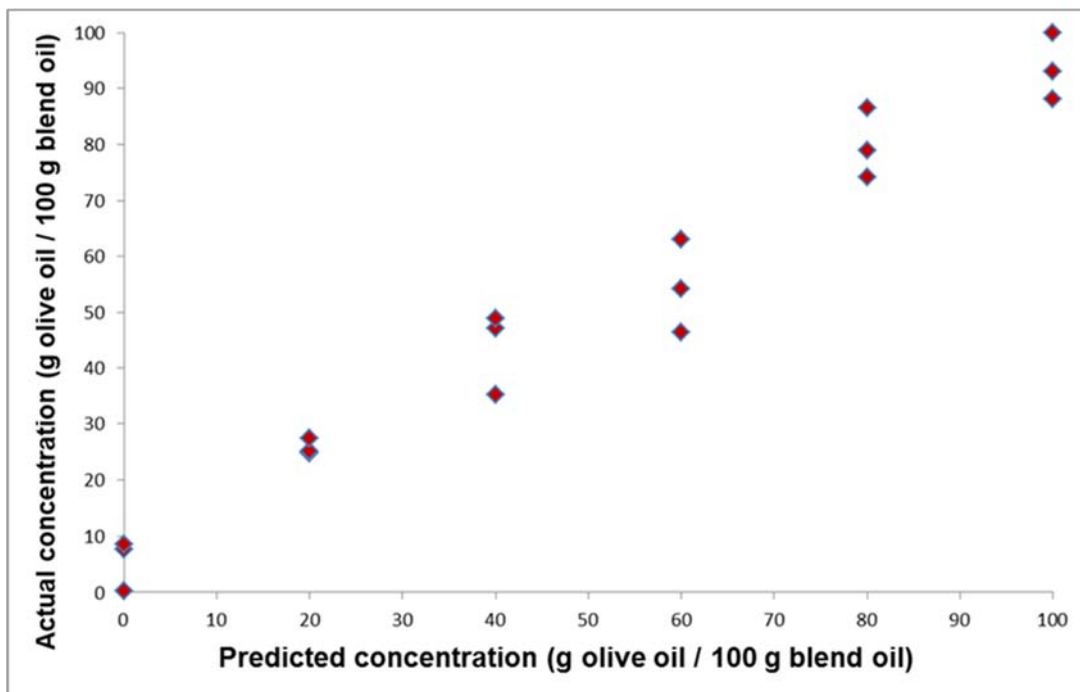
449

Figure 4



450

Figure 5

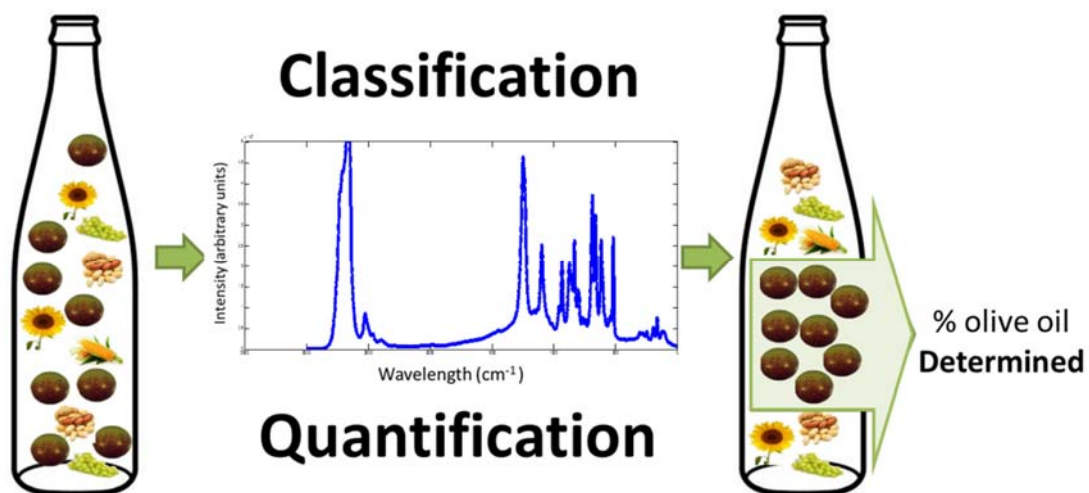


451

452

453

Graphical abstract



454